Journal of Chemical Information and Modeling

Parallel Tempering of Dark Matter from the Ebola Virus Proteome: Comparison of CHARMM36m and CHARMM22 Force Fields with Implicit Solvent and Coarse-Grained Model

Mark A. Olson

Department of Cell Biology and Biochemistry, Molecular and Translational Sciences, USAMRIID, Frederick, Maryland USA

ABSTRACT: Intrinsically disordered proteins are characterized by their large manifold of thermally accessible conformations and their related statistical weights making them an interesting target of simulation studies. To assess the development of a computational framework for modeling this class of proteins, this work examines temperature-based replica exchange simulations to generate a conformational ensemble of a disordered 28-residue peptide from the Ebola virus protein VP35 starting from a prefolded helix-β-turn-helix topology observed in a viral assembly. The simulation strategy tested is the recently refined CHARMM36m force field combined with a generalized Born solvent model to calculate probability density profiles and the results are compared to an equivalent CHARMM22 simulation dataset. The assessment is further extended to include coarse-grained lattice Monte Carlo simulations to determine the accuracy of a reductionism perspective. The analysis finds CHARMM36m to correctly shift the minimum in the potential of mean force to a lower fractional helicity as compared to CHARMM22, however in both simulation models the conformational plasticity along the helix-forming reaction coordinate was limited by free-energy barriers. By comparison the coarse-grained model yielded a potential of mean force of lower resolution as anticipated, yet the model successfully showed multiple equally weighted low-energy states of large-scale conformational heterogeneity not observed in the all-atom models.

1. INTRODUCTION

Intrinsically disordered proteins (IDPs) that encompass "Dark Matter" proteomes play a fundamental role in the regulation and function of protein association networks.[1-4] Their hallmark is large-scale conformational heterogeneity in free solution while finding a folded topology upon forming a multimeric assembly. An illustrative example is a 28-residue peptide region extracted from the Ebola virus protein VP35.[5] The X-ray crystallographic structure of the peptide (designated as NPBP) bound to the Ebola virus protein NP shows a helix-β-turn-helix topology. In free solution NPBP transitions to a disordered ensemble as observed from circular dichroism (CD) spectroscopy.[5] What makes the disorder-order transition of NPBP of larger interest is when added to a solution of 50% trifluoroethanol (TFE) the CD spectrum shows the peptide transitions from an unstructured ensemble to structures containing helical folds. This hidden propensity of NPBP and conceivably that of many other IDPs makes atomistic simulation studies challenging to capture a heterogeneous conformational ensemble without being overly biased by the susceptibility to fold. The challenge is amplified by the inherent deficiencies of all-atom force fields and solvent models that tend to over stabilize fold propensities.

The topic of this brief study is to test the recently reported CHARMM36m force field in a temperature-based replica exchange[6] (T-ReX) simulation of NPBP. The force field is a refinement of an earlier version to improve the accuracy in polypeptide backbone conformational ensembles for IDPs.[7] Although the development of CHARMM36m is primarily intended for explicit solvent simulations, the work here applies the force field with a generalized Born (GB) solvent approximation represented by the GBMV2 model.[8] Given the continued refinement of force fields and optimization studies of GB models,[7,9] it is of general interest to determine if the combined CHARMM36m/GBMV2 simulation strategy provides a framework for modeling IDPs.[10-12] While GBMV2 is a computationally efficient model as compared to explicit solvent simulations and has shown to accurately reconstitute the thermal stability of small α-helical and β-folded proteins,[13-16] questions remain if implicit solvent models can correctly shift the density of states of a prefolded IDP on an energy landscape with multiple kinetic traps to an ensemble of disordered states favored by configurational entropy. To help with the assessment, the CHARMM36m/GBMV2 simulation results are compared to a reassessment of CHARMM22/GBMV2 simulation trajectories taken from an earlier study.[17] The comparison

centers on computing potentials of mean force (PMFs) using the parallel tempering weighted histogram analysis method[18] (PTWHAM) and the multistate Bennett acceptance ratio (MBAR) method.[19]

A further comparison is provided of the all-atom protein simulation models with a coarse-grained (CG) method. The method is based on low-resolution lattice Monte Carlo simulations and explores a reductionism strategy to modeling IDPs. The modeling approach is a revisit of earlier work by Skolnick and coworkers of applying the side-chain-only (SICHO) model[20,21] to NPBP. Accurate reconstruction of all-atom protein representations from CG conformations is applied using the technique developed by Feig and coworkers.[22] Previous successful application of the SICHO model with T-ReX and the rebuilding of all-atom structures is illustrated by the multiscale refinement of protein loops.[23,24] Here, PMFs are calculated from the CG model and compared to the CHARMM36m and CHARMM22 generated conformations.

2. METHODS

CHARMM-based Simulations. For conformational sampling of the 28-residue NPBP peptide, the self-guided Langevin dynamics (SGLD) method developed by Wu and Brooks[25-26] was combined with T-ReX using a strategy first reported in modeling the Trp-cage mini-protein.[14] The simulation methodology and parameter set applied here are similar to that given in an earlier study of the NPBP peptide using the CHARMM22 force field with the GBMV2 solvent model.[17] Here, a summary of the approach is noted of applying CHARMM36m.

The CHARMM simulation program[27] (version c41b1) was applied using the utilities and programming libraries of the Multiscale Modeling Tools for Structural Biology (MMTSB).[28] An integration time step of 2 fs was used and parameters for SGLD consisted of the friction constant set to 1 ps^{-1} for all heavy atoms, the guiding factor λ to a value of 1, and the time average of the momentum was set to 1 ps. Non-bonded interaction cutoff parameters for electrostatics and non-polar terms were set at a radius of 22 Å with a 2-Å potential switching function. Covalent bonds between the heavy atoms and hydrogen atoms were constrained by the SHAKE algorithm.[29] The GBMV2 parameters were selected to smooth the molecular volume by setting *gbmvabeta* = –12 and *gbmvap3* = 0.65.[13] The hydrophobic cavitation term was modeled by applying a phenomenological surface tension coefficient set to a value of 0.015 kcal/mol/$Å^2$.

Simulations were carried out using 24 replica clients and the frequency of exchanges was set to every 1 ps of simulation. Temperatures were geometrically spaced between T_{min} = 300 K and T_{max} = 475 K. The NPBP peptide was modeled for 200 ns of simulation time per client, generating an ensemble of 4.8 μs. Culled structures consisted of 200000 per temperature and were used in the analysis of computing the secondary structure by the DSSP algorithm[30] and the radius of gyration (R_g). PMFs (denoted as the measure W_T at temperature T) were computed using order parameters of fractional helicity (f_H), R_g and Z-score of the potential energies as input to PTWHAM and MBAR calculations. It should be noted that the PMFs generated by the SGLD method are prone to small deviations from a canonical description due to the *ad hoc* force term added to achieve enhanced sampling (see, e.g., Refs. 14 and 26). While an algorithmic scheme to reweight the biases of local averages of forces and momenta has been reported,[26] the application is cumbersome and exceeds the purpose of this work where the deviation is thought to be small for a peptide.[17]

Further analysis of the generated conformations was conducted using MMTSB to evaluate the population density of states by clustering methods that included hierarchical clustering (conformational and f_H values) and *k*-means clustering (conformational). Additional examination was conducted of the root-mean-square deviation (RMSD) in selected backbone angles Φ and Ψ from the initial folded peptide structure. These values were computed across the ensemble for two helical segments and provided input into a PTWHAM calculation.

Lattice Simulations. Chain conformations of NPBP were generated based on Monte Carlo sampling of a cubic lattice using the MONSSTER program developed by Skonick and coworkers.[20,21] The SICHO model was applied where each amino acid is represented by a single virtual particle located at the side-chain center of mass and projected onto a cubic lattice. The force field consists of potential energy terms that account for short- and long-range interactions, hydrogen-bonding cooperativity, and a mean force potential that describes hydrophobic interactions. Each force-field term is constructed of sequence-independent, sequence-dependent, and restraint components. For the purpose of the work presented here, it is noted that the sequence-dependent terms were derived through geometric statistics of known protein structures and account for short-range interactions between nearest neighbors along the polypeptide chain, as well as long-range, pairwise, and soft-core repulsive interactions.[20]

The grid size for the cubic lattice was set at a value of 125 lattice units in each direction at a resolution of 1.45-Å grid spacing. Two separate simulations were conducted, where the parameter *stiff* that controls the scaling of a generic potential term favoring the formation of secondary structure elements was varied. One simulation had the parameter set to a value of 1.0 (where the default is 1.25) and the other set to 0.5. Other parameters were set to their default values given in the MMTSB description. Each of the two simulations was started from the PDB conformation of NPBP and the sequence file annotated with the DSSP secondary structure.

The number of lattice T-ReX simulation cycles at each temperature was set to 20 and the number of Monte Carlo moves per cycle was 50. Culled conformations were extracted from 24 replicas yielding a sampling size comparable to the all-atom simulations. Replicas were exponentially spaced from a reduced temperature T of 1.0 to 2.4, where T is normalized by a reference temperature such that $\beta^{1} = k_{B}T$ represents the energy unit (where k_{B} is the Boltzmann constant). The value $T = 1$ is set to represent the distribution of conformations modeled by the SICHO force field at approximately 300 K. All-atom structures were reconstructed from the lattice simulations by using MMTSB. As with the CHARMM-based simulations, PMFs were calculated and clustering of conformations was conducted.

3. RESULTS and DISCUSSION

The initial conformation of the NPBP peptide (numbered from residues 20 to 47) bound to Ebola virus protein NP is a topology of a helix-β-turn-helix fold and shows a fractional helicity of f_{H} = 0.43 from DSSP. The two helices are partitioned as Trp28 to Thr35 and Val40 to Ile43. The initial fold compactness is given by the dimension R_{g} = 10 Å and appears more collapsed than an estimate for a comparable unfolded 28-residue peptide showing R_{g} ~ 13 Å.[31]

Potentials of Mean Force. Illustrated in Figure 1 are the PMFs at 300 K evaluated by using PTWHAM for the analysis of the CHARMM-based simulations and the CG model. Comparison between CHARMM36m and CHARMM22 shows a clear distinction in population density underlying the conformational ensemble. For CHARMM36m, the minimum in $W_{T}(f_{H},R_{g})$ is observed at f_{H} = 0.26 with a R_{g} = 7.7 Å, while CHARMM22 shows the minimum at f_{H} = 0.53 and R_{g} = 9.8 Å. In addition, CHARMM22 shows a connecting local minimum at $W_{T}(f_{H}$

= 0.37,R_g = 7.9 Å) with a free-energy difference of only 0.05k_BT from the global minimum. To help place the minima in perspective, PTWHAM of conformations taken from an explicit/implicit solvent hybrid T-ReX molecular dynamics simulation study[17] of NPBP with CHARMM22/TIP3P combined with GBMV2 for computing the Metropolis exchanges produced a minimum located at $W_T(f_H = 0.26,R_g = 8.8$ Å). While CHARMM36m with GBMV2 is nearly equal to that observed of the explicit/implicit solvent simulation, the issue is conformational heterogeneity of the generated ensemble.

Overall CHARMM36m corrects the weight of helix propensity compared to CHARMM22, yet the population density of both models remains in disagreement with observations from CD experiments without the addition of TFE. With TFE, the experiments and simulation models are in better agreement in showing preferred helical states, however sampling from CHARMM36m is more confined than CHARMM22 and a possible kinetic trap is revealed. The latter becomes evident in comparing the energy Z-score landscapes, where CHARMM22 simulation shows a manifold of shuttling conformations among the major basins through a pathway of favorable exchanges among the probability distribution. The distinction between the two force fields is also highlighted from the difference in transition from the global minimum in each $W_T(f_H,R_g)$ to an unstructured conformation (f_H = 0, computed at an equivalent R_g as the minimum). The analysis shows for the two force fields that CHARMM36m yields a greater free-energy difference of ~2 k_BT. An advantage for CHARMM36m is that transitions from the minimum to $W_T(f_H = 0,R_g > 10$ Å) shows less frustration than CHARMM22.

To further characterize the landscape topology and explore differences in applying the CHARMM force fields with GBMV2, $W_T(f_H,R_g)$ was recalculated by using the MBAR method. The analysis shows for CHARMM36m an equivalent minimum $W_T(f_H = 0.26,R_g = 8.0$ Å) as calculated from PTWHAM and a slight reweighting of the population density for CHARMM22 to yield $W_T(f_H = 0.46,R_g = 8.1$ Å). One notable difference is observed for CHARMM36m in that MBAR yields a deeper kinetic trap of ~3 k_BT in stabilizing f_H = 0.26 when transitioning to unstructured conformations.

It is important to note that the GBMV2 model is an accurate approximation of the Poisson implicit solvent model[32] and applying the standard protocol of fitted parameters yields the good agreement with explicit solvent simulations of calculating the charging free energy of protein conformations.[33] Deviation to the protocol by modification of Born radii on charged residues and reducing the surface tension of modeling the hydrophobic collapse may lead to lessening the kinetic trap observed in Fig. 1A. Additional improvements may be achieved by incorporating volume and surface area nonpolar solvation free energy terms[34] combined with adaptive parallel tempering methods.[15,17,35] Alternately a unified strategy of re-optimization of the GBMV2 solvent force field for CHARMM36 has been recently proposed by Lee and Chen.[9] While the approach appears promising, further testing is needed for evaluating the adjusted GBMV2 parameters for various applications, including thermal stability of proteins with different fold propensities.

The CG model simulation results of scaling the SICHO potential term favoring the formation of secondary structure elements by the *stiff* parameter of 1 (see Methods) is shown in Fig. 1C of using reconstructed all-atom structures from the lattice generated conformations. PTWHAM of the simulation data is presented and similar results were obtained from MBAR. Analysis reveals the CG model produced significant conformational plasticity among multiple states with free-energy barriers < 0.01 kcal/mol. While $W_T(f_H,R_g)$ is of lower resolution in defining density contours, the CG model exhibits specificity in lattice energies as displayed in the energy Z-score profile. At the same time the model avoids kinetic traps to produce a landscape more consistent with the notion of a disordered ensemble for NPBP. Given the observed conformational heterogeneity, a simple measure of sampled space is the statistical average of populations at $T = 1$ and is given by $f_H = 0.42 \pm 0.14$, where the value approaches the cluster of conformations described by the CHARMM22 $W_T(f_H,R_g)$ minimum. Reducing *stiff* to 0.5 returns an average of 0.11 ± 0.12. In both CG model strategies, the simulations unfold and refold NPBP and selection of parameters near the default values offers a fair assessment of the accuracy in modeling heterogeneity of secondary-structure formation.

Structural Analysis. Displayed in Fig. 2 is the starting NPBP conformation bound to the Ebola virus protein NP. As noted above, NPBP is composed of two helical regions denoted as

α1 of residues Trp28 to Thr35 and α2 for residues Val40 to Ile43. Illustrated in Fig. 2B are conformations taken from clustering the CHARMM36m generated ensemble at 300 K. A further comparison is made with structures from CHARMM22 simulation[17] (Fig. 2C) and a set of all-atom structures reconstructed from the CG generated conformations (Fig. 2D).

Analysis of the extracted structures from CHARMM36m/GBMV2 shows the conformation located at the sampled potential energy minimum (E_{min}) contains a short helix (f_H = 0.26) in the α1 region. This structure best represents the minimum in $W_T(f_H,R_g)$ while the remaining structures exhibit varying helical lengths in α1. Unlike CHARMM36m, the E_{min} of CHARMM22/GBMV2 contains a helix-turn-helix fold of f_H = 0.43 showing similarity to the bound form. For the CG model, contrary to possible chain distortions and entanglements owing to the lower resolution, the simulation yielded well formed structures with E_{min} containing a longer helix of f_H = 0.43. The CG landscape of $W_T(f_H,R_g)$ shows a multiple transitions among the structures where $f_H < \sim 0.5$, while the fold positioned at f_H = 0.75 is of low population.

Figure 2E reports the Cα RMSD of structures extracted from the simulation models at 300 K relative to the prefolded NPBP conformation. The analysis shows the RMSD values are in the range of 5-7 Å of which the CG model reveals the greatest net value. More importantly, the RMSD profiles lack significant disorder of round trips from the initial prefolded state to large Cα excursions. This observation has implications on how the simulation models would represent molecular recognition of NPBP by the protein NP. It suggests the kinetics of recognition is by structural reorganization via "induced-fit" mechanism of helical populations rather than the slow rate-limiting step of capturing a completely unfolded state.

Unlike CHARMM36m, the CHARMM22 and CG models show conformations with helical folds in the α2 region. To further investigate the clustering of structures and the limited folds of a α1-turn-α2 topology found by CHARMM36m, Fig. 3 illustrates PTWHAM assessment of conformational landscapes of RMSD in backbone angles Φ and Ψ from the prefolded peptide for 3-residue segments Ser30-Glu31-Gln32 of α1 and Val40-Ser41-Asp42 of α2. The analysis finds for α1 (Fig. 3A) the highest populated basin to be located at small RMSD differences while the landscape shows considerable conformational populations spread out among the profile. For α2, a strikingly different result is obtained where the landscape shows "hot" population regions

of deviations largely removed from the initial structure, indicating weak stability in the short helix. Interestingly, a further analysis shows that only CHARMM36m/GBMV2 samples a transient β-hairpin in the C-terminal region combined with α1. This arrangement found by CHARMM36m is surprisingly consistent with unbiased predictions of secondary structure by a consensus approach.[36]

Helix Stiffness and Compactness Propensity. Illustrated in Figure 4 are plots of helix formation and fold compactness as a function of sampling temperatures. Shown are statistical averages over datasets for CHARMM36m, CHARMM22 and the CG model. For the latter, effective temperatures from the reduced representation of the lattice simulation model were approximately scaled to those of the all-atom simulations. Also shown in the plots is an assessment of the statistical averages for the CHARMM36m simulation by overlaying values of f_H and R_g where $W_T(f_H,R_g) = 0$ along the temperature profile computed from MBAR.

As anticipated, the analysis shows the CG model to exhibit the weakest cooperativity and thermal stability in helix formation among the simulation models (Fig. 4A). While conformational excursions produced by the CG model sampled significant helical populations, the statistical average is located among multiple low free-energy states where barriers in $W_T(f_H,R_g)$ are ~ 0.01 k_BT. By comparison, the results show CHARMM36m to retain helical states observed in the $W_T(f_H,R_g)$ minima over a 50-K temperature span, before the secondary structure unravels. Of the two all-atom force fields and their thermal profiles, CHARMM36m provides a better model of the experimental CD observations.[5]

One of the concerns of applying implicit solvent descriptions to modeling large-scale conformational heterogeneity is that the sampled structures will be overly collapsed (see, e.g., comparison of solvent models in Ref. 17). Figure 4B shows all three simulation models are plagued by extended compactness of the generated conformations. As a noted benchmark, the prefolded peptide conformation shows a R_g ~ 10 Å in the multimeric complex. A further observation in Fig. 4B reveals significant deficiency in the CG simulation model. While the model samples R_g > 15 Å as shown in the $W_T(f_H,R_g)$, the SICHO force field strongly favors

compacted topologies during replica exchanges, making this simulation approach less attractive for modeling multistate ensembles of thermal unfolding transitions.

4. CONCLUSIONS

This study investigated the application of the CHARMM36m force field with the GBMV2 implicit solvent model in a replica exchange simulation of calculating the conformational ensemble of a 28-residue IDP from the Ebola virus protein VP35. Comparisons were made with data from an equivalent CHARMM22/GBMV2 simulation study and a coarse-grained model of applying a lattice Monte Carlo simulation. The central issue of the study is the applicability of potential energy functions applied in parallel tempering algorithms as a computational approach for modeling large-scale conformational heterogeneity. The measure of success was the accuracy to replicate a disordered conformational ensemble of the peptide as measured from CD experiments. Starting from a helix-turn-helix topology, the results revealed that CHARMM36m combined with GBMV2 produced a potential of mean force of lower fractional helicity than CHARMM22, yet neither simulation model captured significant conformational plasticity along the helix-forming reaction coordinate between unstructured and folded conformations. Moreover, the models displayed a helix propensity with an extended thermal stability over the ensemble and the conformations were overly collapsed in the dimension of radius of gyration. Overall the study demonstrated that the accuracy of the GBMV2 model in its standard protocol with the all-atom force field CHARMM36m is limited in the modeling large-scale conformational heterogeneity of IDPs. The likely best scenario of applying the GBMV2 model with CHARMM36m is the explicit/implicit solvent hybrid replica exchange method where peptide conformations are generated on an explicit solvent landscape. By comparison, the coarse-grained model yielded an ensemble of thermally accessible states showing conformational disorder in the potential of mean force. Like the all-atom models, the lattice-generated conformations were collapsed in fold space on the manifold of highly populated states.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: (M.A.O.)

Notes

The author declared no competing financial interest.

■ ACKNOWLEDGMENTS

Financial support for this work comes from US Department of Defense Threat Reduction Agency. The opinions or assertions contained herein are the private views of the author and are not to be construed as official or as reflecting the views of the US Army or of the US Department of Defense. This article has been approved for public release with unlimited distribution.

■ REFERENCES

(1) Bhowmick, A.; Brookes, D. H.; Yost, S. R.; Dyson, H. J.; Forman-Kay, J. D.; Gunter, D.; Head-Gordon, M.; Hura, G. L.; Pande, V. S.; Wemmer, D. E.; Wright, P. E.; Head-Gordon, T. Finding Our Way in the Dark Proteome. *J. Am. Chem. Soc*. **2016**, *138*, 9730–9742.

(2) Arai, M.; Sugase, K.; Dyson, H. J.; Wright, P. E. Conformational Propensities of Intrinsically Disordered Proteins Influence the Mechanism of Binding and Folding. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 9614–9619.

(3) Wright, P. E.; Dyson, H. J. Intrinsically Unstructured Proteins: Re-Assessing the Protein Structure-Function Paradigm. *J.Mol.Biol*. **1999**, *293*,321–331.

(4) Wright, P. E.; Dyson, H. J. Intrinsically Unstructured Proteins and their Functions. *Nat. Rev. Mol. Cell Biol*. **2005**, *6*,197–208.

(5) Leung, D.W.; Borek, D.; Luthra, P.; Binning, J. M.; Anantpadma, M.; Liu, G.; Harvey, I. B.; Su, Z.; Endlich-Frazier, A.; Pan, J.; Shabman, R. S.; Chiu, W.; Davey, R. A.; Otwinowski, Z.; Basler, C. F.; Amarasinghe, G. K. An Intrinsically Disordered Peptide from Ebola Virus VP35 Controls Viral RNA Synthesis by Modulating Nucleoprotein-RNA Interactions. *Cell Rep*. **2015**, *11*, 376–389.

(6) Ishikawa, Y.; Sugita, Y.; Nishikawa, T.; Okamoto, Y. Ab Initio Replica Exchange Monte Carlo Method for Cluster Studies. *Chem. Phys. Lett*. **2001**, *333*, 199−206.

(7) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. Jr. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat. Methods*. **2017**, *14*, 71–73.

(8) Lee, M. S.; Feig, M.; Salsbury, F. R., Jr.; Brooks, C. L., 3rd. New Analytic Approximation to the Standard Molecular Volume Definition and its Application to Generalized Born Calculations. *J. Comput. Chem*. **2003**, *24*, 1348−1356.

(9) Lee, K. H.; Chen, J. Optimization of the GBMV2 Implicit Solvent Force Field for Accurate Simulation of Protein Conformational Equilibria. *J. Comput. Chem*. **2017**, doi: 10.1002/jcc.24734. [Epub ahead of print].

(10) Best, R. B. Computational and Theoretical Advances in Studies of Intrinsically Disordered Proteins. *Curr. Opin. Struct. Biol*. **2017**, *42*, 147-154.

(11) Chong, S. H.; Chatterjee, P.; Ham, S. Computer Simulations of Intrinsically Disordered Proteins. *Annu. Rev. Phys. Chem*. **2017**, doi: 10.1146/annurev-physchem-052516-050843. [Epub ahead of print].

(12) Bhowmick, A.; Brookes, D. H.; Yost, S. R.; Dyson, H. J.; Forman-Kay, J. D.; Gunter, D.; Head-Gordon, M.; Hura, G. L.; Pande, V. S.; Wemmer, D. E.; Wright. P. E.; Head-Gordon, T. Finding our Way in the Dark Proteome. *J.Am.Chem.Soc*. **2016**, *138*, 9730–9742.

(13) Yeh, I. C.; Lee, M. S.; Olson, M. A. Calculation of Protein Heat Capacity from Replica-Exchange Molecular Dynamics Simulations with Different Implicit Solvent Models. *J. Phys. Chem. B.* **2008**, *112*, 15064-15073.

(14) Lee, M. S.; Olson, M. A. Protein Folding Simulations Combining Self-Guided Langevin Dynamics and Temperature-Based Replica Exchange. *J. Chem. Theory Comput*. **2010**, *6*, 2477−2487.

(15) Lee, M. S.; Olson, M. A. Comparison of Two Adaptive Temperature-Based Replica Exchange Methods Applied to a Sharp Phase Transition of Protein Unfolding-Folding. *J. Chem. Phys*. **2011**, *134*, 244111–224417.

(16) Chaudhury, S.; Olson, M. A.; Tawa, G.; Wallqvist, A.; Lee, M. S. Efficient Conformational Sampling in Explicit Solvent using a Hybrid Replica Exchange Molecular Dynamics Method. *J. Chem. Theory Comput*. **2012**, *8*, 677–687.

(17) Olson, M. A. On the Helix Propensity in Generalized Born Solvent Descriptions of Modeling the Dark Proteome. *Front. Mol. Biosci*. **2017**, *4*, 3. doi: 10.3389/fmolb.2017.00003.

(18) Chodera, J. D.; Swope, W. C.; Pitera, J. W.; Seok, C.; Dill, K. A. Use of the Weighted Histogram Analysis Method for the Analysis of Simulated and Parallel Tempering Simulations. *J. Chem. Theory Comput*. **2007**, *3*, 26–41.

(19) Shirts, M. R.; Chodera, J. D. Statistically Optimal Analysis of Samples from Multiple Equilibrium States. *J. Chem. Phys*. **2008**, *129*, 124105–10.

(20) Skonick, J.; Kolinski, A.; Ortiz, A. R. MONSSTER: A Method for Folding Globular Proteins with a Small Number of Distance Restraints. *J. Mol. Biol*. **1997**, *265*, 217–241.

(21) Kolinski, A.; Skonick, J. Assembly of Protein Structure from Sparse Experimental Data: An Efficient Monte Carlo Model. *Proteins*. **1998**, *32*, 475–494.

(22) Feig, M.; Rothiewiez, P.; Kolinski, A.; Skolnick, J.; Brooks, C. L., 3rd. Accurate Reconstruction of All-Atom Protein Representations from Side-Chain-Based Low-Resolution Models. *Proteins*. **2000**, *41*, 86–97.

(23) Olson, M. A.; Chaudhury, S.; Lee, M. S. Comparison between Self-Guided Langevin Dynamics and Molecular Dynamics Simulations for Structure Refinement of Protein Loop Conformations. *J. Comput. Chem*. **2011**, *32*, 3014–22.

(24) Olson, M. A.; Feig, M.; Brooks, C. L. 3rd. Prediction of Protein Loop Conformations using Multiscale Modeling Methods with Physical Energy Scoring Functions. *J. Comput. Chem*. **2018**, *29*, 820–31.

(25) Wu, X.; Brooks, B. R. Self-Guided Langevin Dynamics Simulation Method. *Chem. Phys. Lett*. **2003**, *381*, 512−518.

(26) Wu, X.; Damjanovic, A.; Brooks, B. R. Efficient and Unbiased Sampling of Biomolecular Systems in the Canonical Ensemble: A Review of Self-Guided Langevin Dynamics. *Adv. Chem. Phys*. **2012**, *150*, 255−326.

(27) Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; *et al*. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem*. **2009**, *30*, 1545−1614.

(28) Feig, M.; Karanicolas, J.; Brooks, C. L., 3rd. MMTSB Tool Set: Enhanced Sampling and Multiscale Modeling Methods for Applications in Structural Biology. *J. Mol. Graphics Modell*. **2004**, *22*, 377−395.

(29) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys*. **1977**, *23*, 327−341.

(30) Kabsch, W.; Sander, C. Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **1983**, *22*, 2577–2637.

(31) Kohn, J. E.; Millett, I. S.; Jacob, J.; Zagrovic, B.; Dillon, T. M.; Cingel, N., et al. Random-coil Behavior and the Dimensions of Chemically Unfolded Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 12491–12496.

(32) Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D. A.; Brooks, C. L. 3rd. Performance Comparison of Generalized Born and Poisson Methods in the Calculation of Electrostatic Solvation Energies for Protein Structures. *J. Comput. Chem*. **2004**, *25*, 265–284.

(33) Lee, M. S.; Olson, M. A. Evaluation of Poisson Solvation Models using a Hybrid Explicit/Implicit Solvent Method. *J. Phys. Chem. B*. **2005**, *24*, 5223–5236.

(34) Lee, M. S.; Olson, M. A. Comparison of Volume and Surface Area Nonpolar Solvation Free Energy Terms for Implicit Solvent Simulations. *J. Chem. Phys.* **2013**, *139*, 044119.

(35) Olson, M. A.; Legler, P. M.; Goldman, E. R.; Comparison of Replica Exchange Simulations of a Kinetically Trapped Protein Conformational State and its Native Form. *J. Phys. Chem. B.* **2016**, *120*, 2234-40.

(36) Kieslich, C. A., Smadbeck, J., Khoury, G. A., Floudas, C. A. conSSert: Consensus SVM Model for Accurate Prediction of Ordered Secondary Structure. *J. Chem. Inf. Model*. **2016**, *56*, 455–461.

■ Figure Legends

Figure 1. Conformational and energy landscapes computed from the three simulation models where the order parameters are radius of gyration, fractional helicity and Z-score of potential energies. Plots display data extracted at T = 300 K for the all-atom simulations and T = 1 for the CG lattice Monte Carlo model. Results are: (**A**) CHARMM36m/GBMV2 simulation results; (**B**) CHARMM22/GBMV2 simulation results; and (**C**) CG simulation. The free energy scale is displayed where the color blue represents minima in the PMFs and is followed in order by colors green, yellow and red, where the latter denotes high energy states of low population.

Figure 2. Analysis of structures extracted from the simulations. (A) The initial prefolded conformation observed in the viral assembly where the small peptide represents NPBP and the molecular surface represents Ebola virus NP. Further conformations are illustrated taken at T = 300 K from the CHARMM36m/GBMV2 simulation. (**B**) Conformations extracted from the CHARMM22/GBMV2 simulation.[17] (**C**) Conformations extracted from the CG model simulation results at T = 1. (**D**) Plots of conformation index versus Cα-RMSD (units of Å) at the lowest temperature from the prefolded conformation where the blue colored line and gray data set represent the CHARMM36m/GBMV2 simulation results, red colored line represents the CHARMM22/GBMV2 simulation and the black colored line represents the CG model.

Figure 3. Landscape of RMSD in Φ and Ψ (units of angle) in modeled conformations from CHARMM36m/GBMV2 simulation compared to the prefolded NPBP for (**A**) residues Ser30-Glu31-Gln32 of α1 and (**B**) residues Val40-Ser41-Asp42 of α2. Colors applied in the PMFs and their scales are noted in Figure 1.

Figure 4. Helix stiffness and compactness propensity of generated conformations by the simulation models along the temperature profile. Line with red color circles represents a statistical average of data from the CHARMM36m/GBMV2 simulation, blue colored line and symbols represent values of f_H and R_g where $W_T(f_H,R_g) = 0$ for CHARMM36m/GBMV2 dataset, long-dashed line represents CHARMM22/GBMV2, and short-dashed line represents the CG model. (**A**) f_H versus T; and (**B**) R_g versus T. Standard deviation for CHARMM36m/GBMV2 simulation generating f_H at T = 300 K is ± 0.10 and ± 0.02 at T = 475 K; for R_g the standard deviation is ± 0.38 Å at T = 300 K and for T = 475 K, deviation is ± 2.64 Å. Error in determining the minima in $W_T(f_H,R_g)$ along the temperature coordinate is ~ 0.1 kcal/mol.

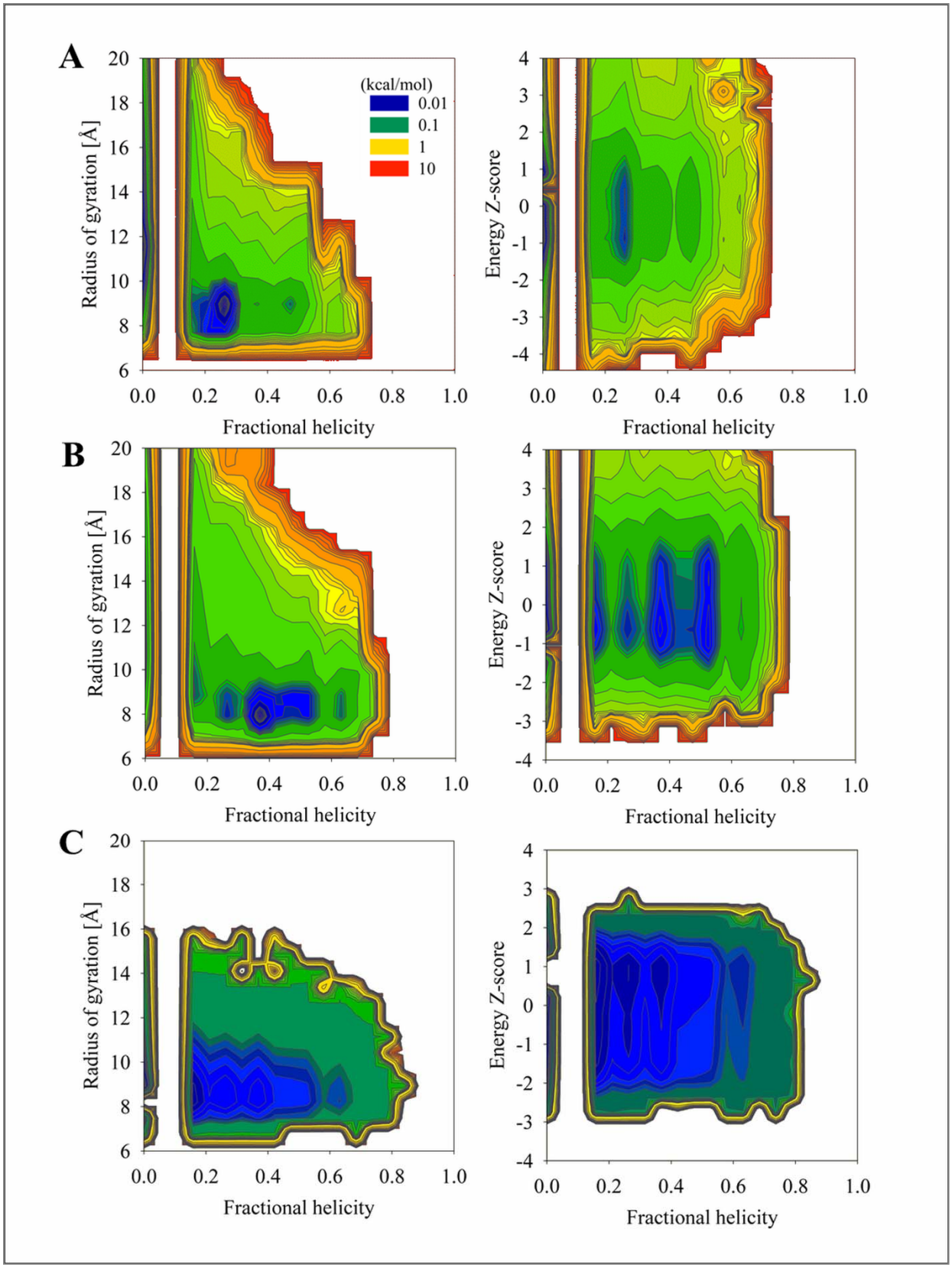
A
B
C
(kcal/mol)
0.01
0.1
1
10
Radius of gyration [Å]
Energy Z-score
Fractional helicity
20
18
16
14
12
10
8
6
4
3
2
1
0
-1
-2
-3
-4
0.0
0.2
0.4
0.6
0.8
1.0

Figure 1

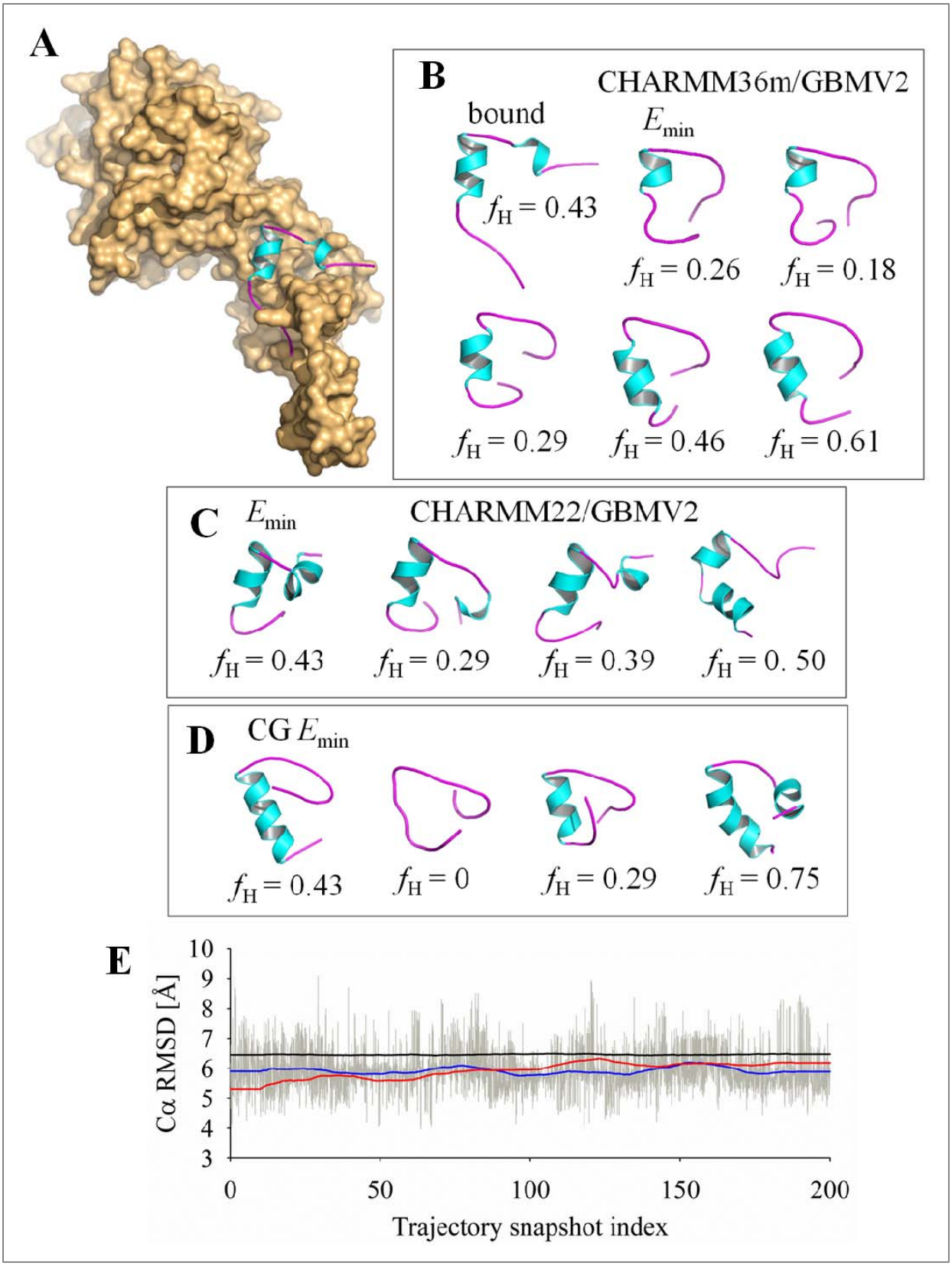
A
B
CHARMM36m/GBMV2
bound
E_{min}
$f_H = 0.43$
$f_H = 0.26$
$f_H = 0.18$
$f_H = 0.29$
$f_H = 0.46$
$f_H = 0.61$
C
E_{min}
CHARMM22/GBMV2
$f_H = 0.43$
$f_H = 0.29$
$f_H = 0.39$
$f_H = 0.50$
D
CG E_{min}
$f_H = 0.43$
$f_H = 0$
$f_H = 0.29$
$f_H = 0.75$
E
Cα RMSD [Å]
10
9
8
7
6
5
4
3
0
50
100
150
200
Trajectory snapshot index

Figure 2

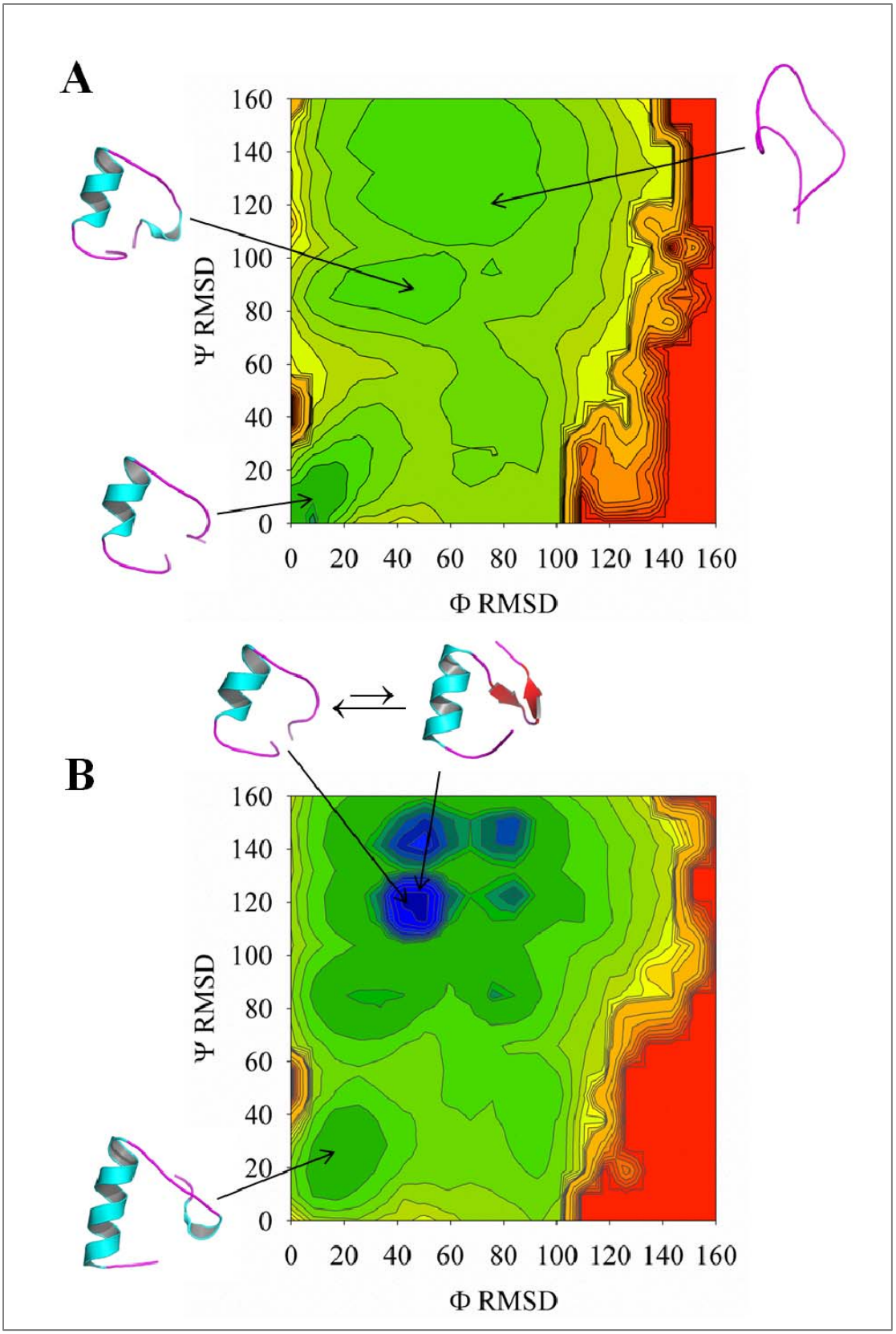
A
160
140
120
100
80
60
40
20
0
Ψ RMSD
0 20 40 60 80 100 120 140 160
Φ RMSD
B
160
140
120
100
80
60
40
20
0
Ψ RMSD
0 20 40 60 80 100 120 140 160
Φ RMSD

Figure 3

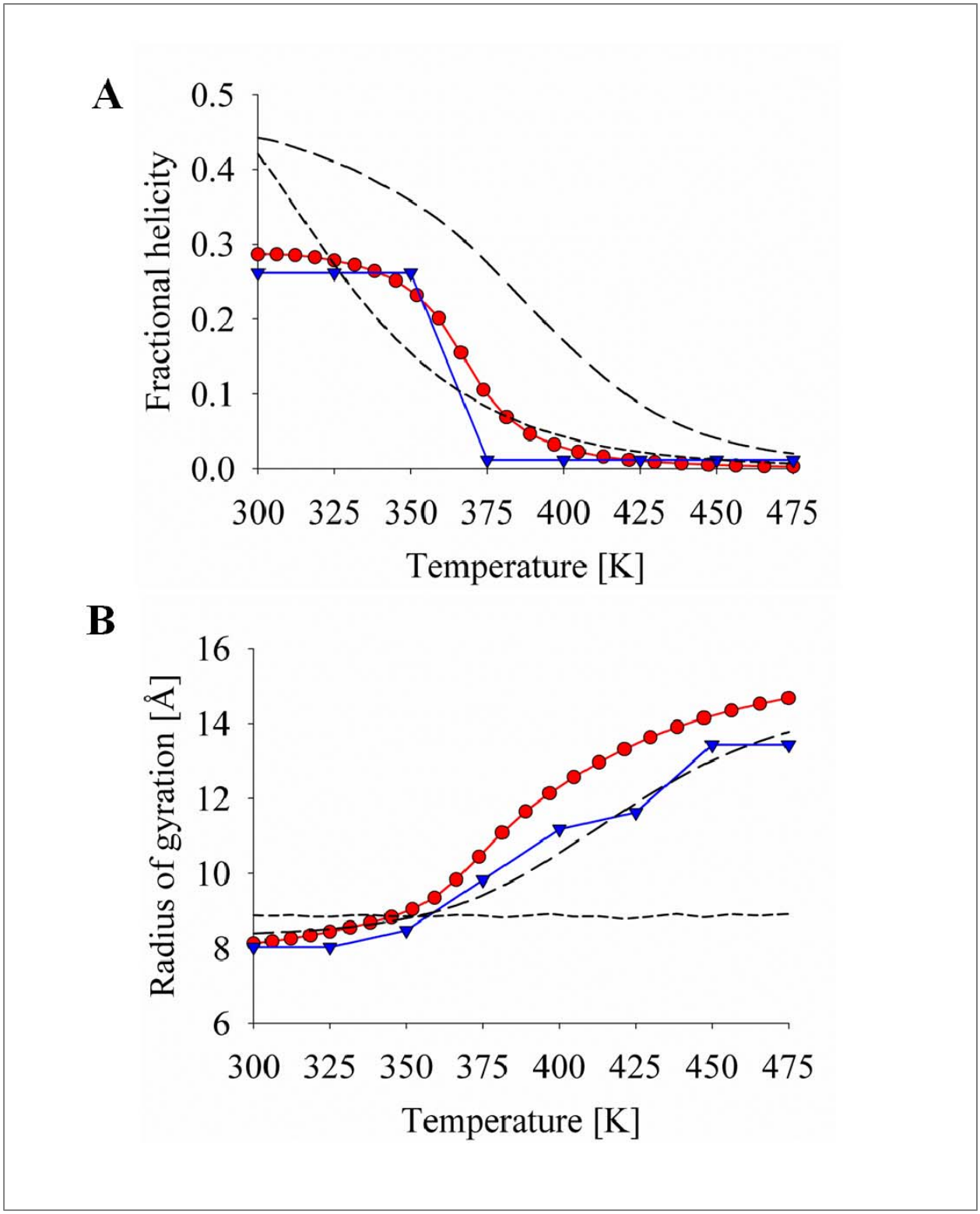
A
Fractional helicity
0.5
0.4
0.3
0.2
0.1
0.0
300 325 350 375 400 425 450 475
Temperature [K]
B
Radius of gyration [Å]
16
14
12
10
8
6
300 325 350 375 400 425 450 475
Temperature [K]

Figure 4

Journal of Chemical Information and Modeling

Table of Contents Graphics